| Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503. |

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED RePrint Report 2004-2005
4. TITLE AND SUBTITLE Eye movements and visual memory for scenes		**5. FUNDING NUMBERS** W911NF0410078
6. AUTHOR(S) John M. Henderson and Monica S. Castelhano		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Contract and Grant Administration Michigan State University 301 Administration Bldg East Lansing MI 48824-1046		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211		10. SPONSORING / MONITORING AGENCY REPORT NUMBER 45832.11-LS

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In this chapter we discuss three types of memory that are relevant for understanding how scene representations are generated over the course of scene viewing. We focus particularly on scene memory generated dynamically across eye movements, and we highlight studies that record eye movements. We argue that the results of studies focusing on transsaccadic memory, active on-line scene memory, and long-term scene memory converge on the conclusion that relatively detailed visual scene representations are retained both over the short and long term, and that these representations are generated incidentally as a consequence of scene viewing.

14. SUBJECT TERMS Scene perception, eye movements, real-world scenes, visual attention.			15. NUMBER OF PAGES 14
			16. PRICE CODE
17. SECURITY CLASSIFICATION OR REPORT **UNCLASSIFIED**	18. SECURITY CLASSIFICATION ON THIS PAGE **UNCLASSIFIED**	19. SECURITY CLASSIFICATION OF ABSTRACT **UNCLASSIFIED**	20. LIMITATION OF ABSTRACT UL

Enclosure 1

9

Eye movements and visual memory for scenes

John M. Henderson and Monica S. Castelhano

Abstract

In this chapter we discuss three types of memory that are relevant for understanding how scene representations are generated over the course of scene viewing. We focus particularly on scene memory generated dynamically across eye movements, and we highlight studies that record eye movements. We argue that the results of studies focusing on transsaccadic memory, active on-line scene memory, and long-term scene memory converge on the conclusion that relatively detailed visual scene representations are retained both over the short and long term, and that these representations are generated incidentally as a consequence of scene viewing.

Introduction

During natural scene viewing, the eyes move to a new fixation location about three times each second (Henderson and Hollingworth 1998; Henderson 2003; see Fig. 9.1), yet we do not experience the tens of milliseconds that transpire during the saccadic movements as blank periods or 'holes' in our visual experience, nor do we experience the visual world as the series of discrete snapshots. Instead, we have the perceptual experience of a complete, full color, highly detailed, and stable visual world. That is, our perceptual experience suggests to us that the visual system in some sense creates a high-resolution internal copy of the external world. Indeed, this phenomenology has historically motivated much of the theoretical work in human and computer vision, and the experience of a complete and detailed visual world has been a major consideration in recent theoretical treatments of scene representation, visual memory, and the nature of consciousness (e.g. Dennett 1991; O'Regan 1992; Rensink 2000a; Wolfe 1999).

Reductions in visual acuity and color sensitivity as a function of distance from the center of fixation place severe constraints on the generation of a detailed internal visual representation of the external scene, so creation of such a representation would require the storage of visual information across each saccade, with representations from consecutive fixations integrated in some way. Furthermore, such representations
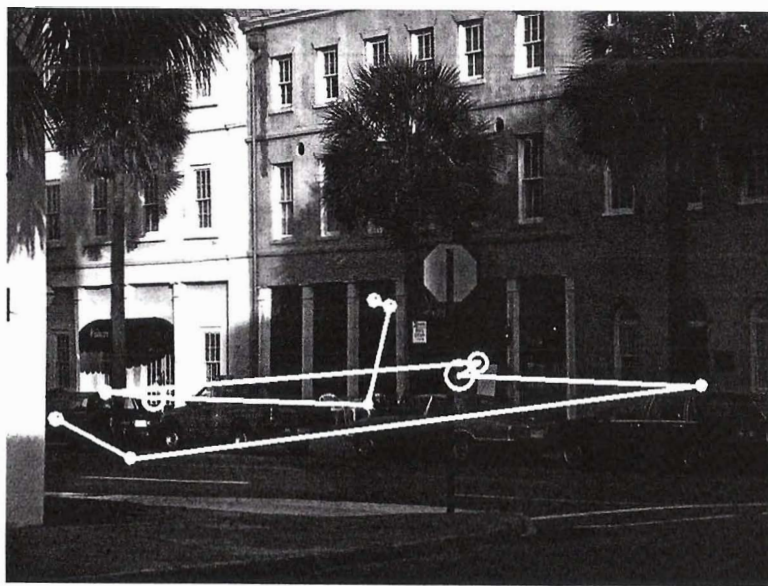
Figure 9.1 During scene viewing, the eyes move to a new fixation location about three times per second on average. In this figure, a participant was viewing the scene while searching for people. Lines represent saccades and circles represent fixations (circle size is scaled to fixation duration). Note that the original images were presented in color.

would have to be retained in active on-line memory over multiple-fixation saccade cycles if they were to be integrated over the entire course of scene viewing. Finally, once constructed, such representations would need to be stored in longer-term memory so that they would be available to support future viewing, perceptual learning, and other cognitive activities such as visual thinking and reasoning, as well as language use (see Henderson and Ferreira 2004). In the following sections we briefly review the evidence for retention and integration of visual representations over a single saccade (transsaccadic memory), over multiple-fixation saccade cycles (active on-line scene memory), and over the longer term (long-term scene memory). We use these categories as an expository device to help organize the literature, and make no claim that retention and integration over these different time scales requires separate structural memory stores. We do not attempt an exhaustive review, but rather try to highlight some of the critical studies as we see them, with an emphasis on eye-movement research and specifically on recent experiments from our laboratory. Our conclusion is that relatively detailed (though not sensory or iconic) visual representations are generated and retained in memory as a natural consequence of active, dynamic scene perception.

Transsaccadic memory

What is the nature of the representation that is retained and integrated across saccades? A proposal with a venerable history is that high-resolution sensory images are stored across saccades, with images from consecutive fixations integrated to form a composite sensory image (for reviews see Bridgeman et al. 1994; McConkie and Currie 1996). Traditionally, this spatiotopic fusion hypothesis (Irwin 1992a) has been instantiated by models in which a sensory image (i.e. a precise, highly detailed, metrically organized, pre-categorical image) is generated during each fixation and stored in a temporary buffer, with sensory images from consecutive fixations spatially aligned and fused in a system that maps a retinal reference frame onto a spatiotopic frame (Breitmeyer et al. 1982; Davidson et al. 1973; Duhamel et al. 1992; Feldman 1985; Jonides et al. 1982; McConkie and Rayner 1976; O'Regan and Lévy-Schoen 1983; Pouget et al. 1993; Trehub 1977). In such models, the composite image formed during consecutive fixations is aligned by tracking the extent of the saccade (via afferent or efferent pathways) or by comparing the similarity of the individual images.

Although many versions of the sensory fusion hypothesis have been proposed, the vast majority of psychophysical and behavioral evidence from the vision and cognition literatures has failed to support it. Perhaps the most convincing evidence arises from direct demonstrations that viewers are unable to fuse simple visual patterns across saccades. In these studies, viewers are required to integrate a pre-saccade and post-saccade pattern in order to accomplish the task successfully. If visual patterns can be fused in a spatiotopically-based sensory memory system, then performance should be similar in a transsaccadic condition in which the environmental spatial position of the patterns is maintained but retinal position is displaced due to a saccade, and a condition in which position in both retinal and environmental spatial reference frames is maintained within a fixation. For example, when two dot patterns forming a matrix of dots are presented in rapid succession at the same retinal and spatial position within an eye fixation, a single fused pattern is perceived and performance (e.g. identification of a missing dot from the matrix) can be based upon this percept (Di Lollo 1980; Eriksen and Collins 1967; Irwin 1991). However, when the two patterns are viewed with similar timing parameters at the same external spatial position but different retinal positions across a saccade, no such fused percept is experienced and performance is dramatically reduced (Bridgeman and Mayer 1983; Irwin 1991; Irwin et al. 1988, 1983, 1990; Jonides et al. 1983; O'Regan and Lévy-Schoen 1983; Rayner and Pollatsek 1983). In the latter case, overall performance is limited to and constrained by the capacity of short-term memory (Irwin et al. 1988). Other effects which might be expected based on the formation of a composite image via sensory fusion, such as spatiotopically-based visual masking, are also not observed (Irwin et al. 1988; Irwin et al. 1990). For other reviews of this work, see Irwin (1992b), Irwin and Andrews (1996), Pollatsek and Rayner (1992), and Rayner (1998).

If the visual information acquired from successive fixations is fused into a single composite sensory image, then displacements of the viewed world during a saccade should be highly noticeable and troublesome because fusion should be disrupted. Contrary to this prediction, Bridgeman et al. (1975) demonstrated that a scene could be spatially

displaced during a saccade with no conscious experience that the stimulus had shifted position, and with little or no disruption to the performance of a visual task. This insensitivity to spatial displacement across saccades has subsequently been replicated many times (e.g. Bridgeman and Stark 1979; Currie *et al.* 2000; Henderson 1997; Irwin 1991; Mack 1970; McConkie and Currie 1996; Verfaillie *et al.* 1994; Whipple and Wallach 1978). An interesting exception to these findings was reported by Deubel *et al.* (1996, 2002; Gysen *et al.* 2002). In these experiments, participants were found to be sensitive to spatial displacements of a target during a saccade when a blank interval was inserted following the saccade and prior to the reappearance of the spatially shifted target. This is an intriguing finding regarding the retention of information across saccades and suggests that sensory memóry may persist during the saccade. However, given that there is typically no blank period at the beginning of each new fixation, it is not clear how this retained information would be functional in the transsaccadic integration process.

Changes to other visual properties are similarly difficult to detect across a saccade. For example, readers are insensitive to changes in the visual properties of text from fixation to fixation (McConkie and Zola 1979). In these experiments, participants read text made up of characters of alternating case. During a given saccade, the case of all characters was exchanged. These case changes were not noticed by readers and had very little if any effect on reading rate or comprehension. Similar insensitivity to changes in visual features of an image across a saccade has been shown with pictures of objects and scenes. For example, Henderson (1997) found that it was very difficult for observers to detect a change to the specific contours of an object from fixation to fixation (see Fig. 9.2). In this study, participants were asked to fixate a point on a computer screen. A line drawing of an object was then presented to the right of fixation. About half of the contours of the object were presented; the other contours were occluded by black stripes. The participant executed a saccade to the object as soon as it appeared. During the saccade, the object remained exactly the same; changed to reveal the complementary set of contours; shifted one stripe width in position; or changed to a different object. The participant was asked to indicate if any change occurred. Participants failed to detect the majority of contour changes or position shifts. In a control condition in which the changes took place at the same retinal and spatial position (and at the same visual eccentricity as the preview had appeared) within a fixation, change detection was quite good. This latter result ensured that the contours and positions could be discriminated at the visual eccentricity used in the transsaccadic change experiment. Henderson and Hollingworth (2003a) reported similar results for full scenes. Other visual changes such as enlargements and reductions of object size often go unnoticed when they take place during a saccade (Henderson *et al.* 1987; Pollatsek *et al.* 1984).

What is retained across saccades?

Irwin and colleagues have demonstrated in a transsaccadic partial report task that the perceptual properties of up to four visual patterns can be retained across saccades in visual short-term memory (Irwin and Andrews 1996). Carlson-Radvansky (Carlson-Radvansky 1999; Carlson-Radvansky and Irwin 1995) found that structural descriptions of simple visual patterns can be retained across saccades. In transsaccadic
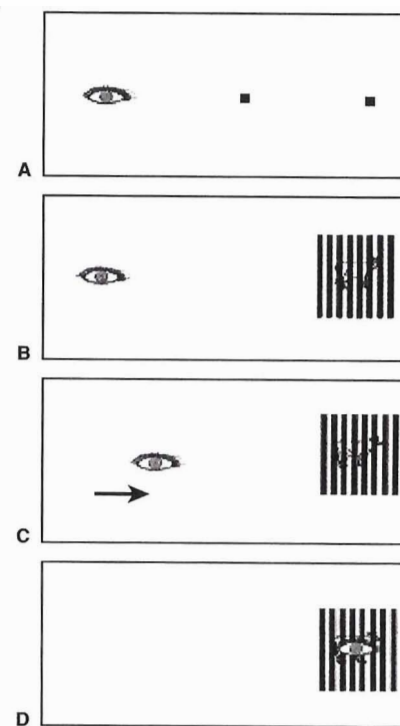

Figure 9.2 Illustration of the study reported by Henderson (1997). Participants began by fixating a point (Panel A). An image of a line drawing of an object was presented to the right of fixation, with about half of the contours occluded by black stripes (Panel B). The participant executed a saccade to the object, and during the saccade the object remained the same, changed to reveal the complementary set of contours, shifted one stripe width in position, or changed to a different object (Panel C). Following fixation, the participant indicated if any change occurred or named the object (Panel D). Contour changes and position shifts were very difficult to detect and did not affect naming latencies.

object identification studies, participants are quicker to identify an object when a preview of the object is available prior to the saccade than when no preview is available (e.g. Henderson 1992a, 1994, 1997; Henderson and Siefert 1999, 2001; Henderson *et al.* 1987, 1989; Pollatsek *et al.* 1984, 1990). Furthermore, preview benefits for objects can be affected by visual changes such as replacement of one visual token with another token of the same conceptual type (Henderson and Siefert 2001) and mirror reflections (Henderson and Siefert 1999, 2001). The influence of visual change on preview benefit is more pronounced when the spatial location of the target object remains constant compared with when the location changes (Henderson 1994; Henderson and Anes 1994; Henderson and Siefert 2001).

The transsaccadic integration results strongly suggest that visual properties are preserved in the representations that are retained across saccades, and furthermore that such representations are at least partially tied to spatial position. It is important to note, however, that visual representations need not be sensory. That is, representation of detailed visual information does not imply the preservation of an iconic image. For example, in the study reported in Henderson (1997) and shown in Fig. 9.2, replacement of the specific contours present in the image from preview to fixation had no effect on preview benefit as assessed by naming latency. We take sensory representation to refer to a complete, precise, pre-categorical, maskable, and metrically organized image of the visual scene (Irwin 1992b; Neisser 1967; Sperling 1960). In contrast, a post-sensory visual representation is an imprecise, post-categorical, non-maskable, and non-iconic visual description encoded in the vocabulary of visual computation. This same distinction maps onto the distinction in the 'iconic memory' literature between visual and informational persistence on the one hand, and visual short-term memory on the other (Irwin 1992b; see also Coltheart 1980). Based on explorations of integration of visual patterns across saccades, Irwin has argued that transsaccadic memory is in fact visual short-term memory (e.g. Irwin 1991, 1992b). Importantly, however, abstract visual representations are still visual in the sense that they represent visual properties such as object shape and viewpoint, albeit in a non-sensory format. An example of a non-sensory representation of shape is a structural description; as noted above, recent evidence suggests that shape may be encoded and retained across saccades in this representational format (Carlson-Radvansky and Irwin 1995; Carlson-Radvansky 1999). In our view, abstract visual representations are neither sensory, nor are they equivalent to conceptual representations (which encode semantic properties) or linguistic descriptions. Examples of abstract visual representations are structural descriptions (e.g. Biederman 1987; Marr 1982; Palmer 1977) and hierarchical feature representations (e.g. Riesenhuber and Poggio 1999).

Active memory: On-line scene representations

Transsaccadic memory as it is traditionally studied concerns the retention of scene information for the very short period of time that transpires from one fixation to the next during saccadic eye movements – durations typically on the order of 20–80 ms. In this section we consider on-line scene representations that are kept active in visual (working) memory over the course of the current perceptual episode lasting several seconds and multiple-fixation saccade cycles. Much of the recent change detection research in scene perception has tapped into active scene memory in this sense. For example, in a classic initial demonstration of 'change blindness', McConkie and Grimes (McConkie 1990, 1991; see Grimes 1996) had viewers study full-color photographs of scenes in preparation for a relatively difficult memory test. The viewers were told that something in a scene might occasionally change and that they should press a button if and when they noticed such a change. Eye movements were monitored with a dual-Purkinje-image eyetracker so that part of each scene could be changed quickly during a saccade. The changes took place during the *n*th saccade, where *n* was predetermined prior to the onset of the scene. The decisive result was that viewers often failed to detect what would

seem to be very obvious visual changes. For example, none of the participants detected that the hats on two central men in a scene switched heads (Grimes, 1996). Unlike the transsaccadic integration experiments described in the preceding section, scene viewing took place over multiple fixations both before and after the change, so the opportunity was available for constructing an on-line representation over extended viewing time both prior to and after the change.

Reduced sensitivity to visual changes in scenes across saccades has been shown for changes to spatial orientation, color, and object presence (Grimes, 1996; Henderson and Hollingworth 1999a; McConkie 1991; McConkie and Currie 1996). Reduced sensitivity is also observed in paradigms that simulate saccades, such as when a blank field is inserted between two scene images (Rensink *et al.* 1997). In general, it appears that when the local transient motion signals that usually accompany a visual change are unavailable, as is the case during a saccade, sensitivity to what would otherwise be a highly-visible change becomes reduced, and in the extreme case eliminated. These results were initially taken to call into question the view that a detailed visual scene representation is constructed on-line in memory during scene viewing (e.g. O'Regan 1992; Rensink 2000a, 2000b; Wolfe 1999).

In the past few years it has become clear that the original interpretation of 'change blindness' was incorrect and that relatively detailed on-line visual memory representations of scenes can be observed in change detection experiments. Two general sources of evidence converge on this conclusion (for more extensive review, see Henderson and Hollingworth 2003b). In one set of experiments, a change detection paradigm was used in which a target object was changed during a saccade within the scene (Henderson and Hollingworth 1999a; Hollingworth and Henderson 2002; Hollingworth *et al.* 2001). As in the original saccade-contingent scene change experiments (McConkie 1990, 1991; see Grimes 1996) participants viewed pictures of scenes to prepare for a difficult memory test, and in addition were asked to monitor for changes. A target object changed during a saccade toward the object (toward condition); away from the object after it had been fixated the first time (away condition); or during a saccade to a different non-target object elsewhere in the scene (other object condition). In several experiments, the change to the other object region was only activated after the object had received at least one fixation (e.g. Hollingworth and Henderson 2002; Hollingworth *et al.* 2001). Thus, in the away and other object conditions, the target object was attended at some point prior to the scene change but visual attention was directed away from the target object and to a different object in the visual field prior to the saccade that triggered the change (visual attention must be allocated to the target of the next saccadic eye movement prior to the saccade's execution, Henderson 1992b, 1993, 1996; Hoffman and Subramanian 1995; Kowler *et al.* 1995; Shepherd *et al.* 1986). A viewer's ability to detect changes in these conditions provides evidence about whether visual object representations are preserved after the withdrawal of attention and can accumulate during extended visual exploration of a scene.

Four change manipulations have been tested using the on-line saccade-contingent change paradigm: deletion of the target object from the scene (Henderson and Hollingworth 1999a; 2003c); type changes in which the target object is replaced with another object from a different basic-level category (Henderson and Hollingworth 2003c;

Hollingworth and Henderson 2002); token changes in which the target object is replaced with another object from the same basic-level category (Henderson and Hollingworth 2003c; Hollingworth and Henderson 2002; Hollingworth *et al.* 2001), and rotations in which the target object is rotated 90° around its vertical axis (Henderson and Hollingworth 1999a; Hollingworth and Henderson 2002). In each of these conditions, change detection for previously attended objects was significantly above the false alarm rate (which is typically very low). These results suggest that a memory representation is generated and available during on-line scene viewing. It is possible that deletions and type changes could be detected based on semantic information (e.g. deletions and type changes could alter scene meaning), but token changes and rotations as implemented in these experiments did not alter the gist of the scenes in which they appeared.

In the studies described above, object changes were sometimes not detected in the away and other object conditions when they first occurred, but were then detected when the changed object was fixated later in the course of scene viewing. Viewers typically fixated many intervening scene regions and objects over the course of the several seconds that transpired between the initial fixation on the target, the target change, and the first refixation of the (now changed) target object in these cases. The observed delayed change detection following target refixation establishes that the on-line visual scene representation survived over time and potential interference from other fixated objects, and suggests that refixating the changed object provided a cue to retrieve and compare the stored visual object representation to current perceptual information. Hollingworth (2003) has provided additional evidence supporting the hypothesis that change detection failure is often a retrieval problem.

A second source of evidence that change detection in the saccade-contingent change experiments is based on an on-line memory system that lasts longer than a single saccade comes from a manipulation of the semantic consistency of the target object in the scene (Hollingworth *et al.* 2001). In this study, participants viewed line drawings of scenes in which the target object was either semantically consistent or inconsistent with the scene. The target object was replaced by another token of the same basic-level category (e.g. one chicken was replaced by a different chicken in a farm scene) during a saccade away from that object. Scene memory research has demonstrated that the memory representation of a semantically inconsistent object in a scene is more detailed and/or complete compared with a semantically consistent object (e.g. Friedman 1979). Based on these prior results, if visual representations accumulate on-line in memory during scene viewing, then changes to semantically inconsistent objects (which should be represented more completely) should be detected more accurately than changes to semantically consistent objects. The results confirmed this prediction. Furthermore, because the change occurred during the saccade away from the target object, the change was not always detected immediately (see also Henderson and Hollingworth 1999b; Hollingworth and Henderson 2002). Overall, 41 per cent of the detection responses took place more than 1.5 s after the change, and of these responses, 94 per cent occurred only after the target object was refixated. Again, these results suggest that visual object representations are maintained on-line over the course of scene viewing and across extended time and multiple-fixation saccade cycles.

The use of change detection to study the nature of the representations generated during scene perception assumes that the experience of change directly reflects the underlying representation. However, contrary to this assumption, we have found that overt change detection often significantly underestimates the degree to which on-line visual representations are retained in memory. Specifically, gaze duration (the sum of the durations of all fixations from the initial fixation on an object region to the first saccade taking the eyes away from that object) is elevated for trials in which a change occurred but was not reported, compared with no-change control trials (Henderson and Hollingworth 2003c; Hollingworth and Henderson 2002; Hollingworth *et al.* 2001). For example, in one study we found that when a token change was not explicitly detected, mean gaze duration on that object after the change was 749 ms, whereas mean gaze duration was 499 ms when no change occurred (Hollingworth *et al.* 2001). As found for delayed explicit detection, this 'implicit' or covert detection effect was observed despite several intervening seconds and many fixations on other objects between the object change and the first refixation of the target (when gaze durations were found to be elevated). These results provide strong evidence that visual object representations were available in on-line scene memory over the course of viewing even when they were not easily reportable. Thus, the failure to report a change does not provide unambiguous evidence that the information needed to detect that change is unavailable in on-line memory (see also Fernandez-Duque and Thornton 2000; Hayhoe *et al.* 1998; Williams and Simons 2000). In summary, despite the failure of participants to report scene changes taking place during saccades, it is clear that when fixation duration is used to assess the underlying visual representation, robust evidence of visual representation is obtained.

Given the potential difficulty of interpreting overt change detection failure, Andrew Hollingworth (Hollingworth and Henderson 2002) developed a forced-choice memory test to directly investigate viewers' on-line memory for objects in scenes (see Fig. 9.3).
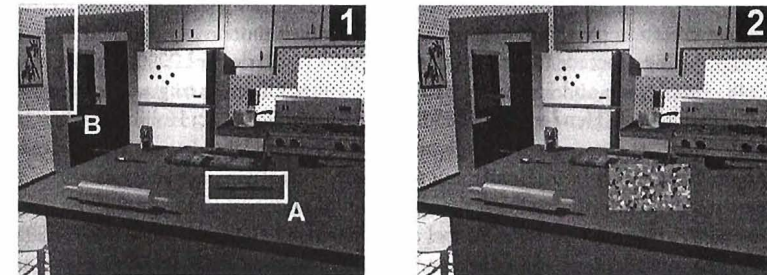

Figure 9.3 Illustration of the paradigm developed by Hollingworth and Henderson (2002). Participants freely viewed each scene, and after the target object (region A) had been fixated, a saccade to another pre-defined object in the scene (region B) triggered masking of the target object. A forced-choice memory test was then presented for the target object. Note that the boxes surrounding regions A and B were not visible to the participants, and the original images were presented in color.

Participants viewed images of common environments while their eye movements were monitored. Following the start of each trial, the computer waited until the target object (indicated by Region A in Fig. 9.3) had been fixated at least once, assuring that it had been attended prior to the test. Then, during a saccade to another object on the other side of the scene (Region B in Fig. 9.3), the target object was obscured by a pattern mask. The onset of the mask coincided with a saccade to a different object in the scene, so the target object was not attended at the time the mask appeared. Following the appearance of the mask, a forced-choice memory test was presented in which two object alternatives were displayed sequentially within the scene: the original target and a distractor object. The distractor was either a different token from the same basic-level category (token discrimination) or a version of the target object rotated 90° in depth around the vertical axis (orientation discrimination).

Performance in this memory test was very good: token discrimination was 87 per cent correct and orientation discrimination was 82 per cent correct. Again, on many trials viewers fixated multiple objects between the last fixation on the target object and the onset of the mask (and the initiation of the forced-choice test), but performance did not differ statistically as a function of the number of intervening fixations; when nine or more fixations intervened between the last fixation on the target object and the onset of the memory test, performance in the token discrimination test was 85 per cent correct and performance in the orientation discrimination test was 92 per cent correct. These results suggest that on-line scene representations are relatively stable in memory. These data, along with the change detection results reviewed above, provide very strong evidence that visual representations from previously attended objects accumulate on-line in memory, forming a relatively detailed scene representation.

Contrary to proposals based on change blindness, visual object representations are not lost upon the withdrawal of attention. At the same time, change blindness clearly is mediated by attention, presumably because attention is needed to encode the pre-change and post-change regions as well as to facilitate retrieval of the pre-change representation from memory following the change. In the transsaccadic change detection paradigm, the same change is much more easily detected when it occurs during a saccade toward the changing object (Currie *et al.* 2000; Hayhoe *et al.* 1998; Henderson and Hollingworth 1999a, 2003c) than during a saccade away from that object (Henderson and Hollingworth 1999a, 2003c; Hollingworth *et al.* 2001). Similarly, transsaccadic integration is heavily weighted toward the saccade target (Henderson 1994; Henderson and Anes 1994; Irwin and Andrews 1996), at least partly due to the fact that the allocation of attention to the saccade target is mandatory prior to a saccade (Deubel and Schneider 1996; Henderson 1992b, 1993, 1996; Henderson *et al.* 1989; Hoffman and Subramanian 1995; Irwin and Gordon 1998; Kowler *et al.* 1995; Rayner *et al.* 1978; Shepherd *et al.* 1986). In the change blindness literature, detection of change is better in the flicker paradigm for scene regions rated to be of higher interest (Rensink *et al.* 1997), for semantically unexpected objects (Hollingworth and Henderson 2000), at locations to which attention has been explicitly directed (Scholl 2000), and at locations near fixation (Hollingworth *et al.* 2001).

Long-term scene memory

In this section, we briefly consider the evidence concerning eye movements, scene representations, and long-term memory. We take long-term scene memory to involve the representations that linger once the current perceptual episode is over. For example, if the current perceptual episode involves working at your desk, in which an active visual representation of the desktop may be generated and maintained, what continues to reside in memory about your desk if you go to another room to watch the Red Sox win the World Series on television? This issue is typically operationalized by studying long-term memory for pictures of scenes.

Classic scene memory research has demonstrated very good long-term memory for scene detail. For example, Nickerson (1965) had participants view 200 black-and-white photographs for 5 s each; on an old-new recognition test, participants correctly recognized 92 per cent of the pictures (controlling for false alarm rate). Shepard (1967) similarly demonstrated 97 per cent correct recognition for 612 color pictures when tested immediately and 99.7 per cent when tested 2 h later. Standing *et al.* (1970) showed participants 2560 pictures for 10 s each over several days. Memory for the entire set of pictures was well over 90 per cent correct. Furthermore, memory for a subset of 280 thematically similar scenes, which required remembering more details about the pictures than general category or gist, was 90 per cent correct. Standing *et al.* (1970) also manipulated the left–right orientation of the scene at study and test, and showed that participants could recognize the studied picture orientation 86 per cent of the time after a 30 s retention interval and 72 per cent of the time after 24 h.

Hollingworth and Henderson (2002) tested long-term memory for individual objects in scenes. A difficult forced-choice discrimination memory test was given for specific target objects after the scenes were removed from view for between 5–30 min. Similar to the on-line memory test described in the previous section, for each studied scene, participants viewed two versions of the scene in the test session: one that was identical to the studied scene and a distractor scene that differed only in the target object. The distractor object was a different type, different token, or the same object rotated in depth. This longer retention interval did not cause a significant decrement in discrimination performance compared with online discrimination. Mean type-discrimination performance was 93 per cent correct, mean token-discrimination performance was 80.6 per cent correct, and mean orientation-discrimination performance was 82 per cent correct. The similarity between discrimination performance in the on-line and long-term tests suggests that visual object representations are stable after attention is removed, at least over the retention intervals we tested. These long-term memory results are consistent with evidence from the picture memory literature cited above suggesting very good memory for the visual form of whole scenes (Standing *et al.* 1970) and for the visual form of individual objects within scenes (Friedman 1979; Parker 1978).

Results from other scene memory studies also support the notion that visual details of objects are encoded in memory (Bahrick and Boucher 1968; Mandler and Ritchey 1977; Mandler and Parker 1976). Although relatively simpler 'scene sketches' (Henderson and Ferreira 2004) were used in these earlier studies (line drawings

of 6–9 objects in each), participants were able to distinguish between the target object and similar distractors of the same basic-level category (Bahrick and Boucher 1968) and were able to recall different types of visual details (Mandler and Ritchey 1977). Several studies have also demonstrated that over the long term, participants are able to recognize object types (Goodman 1980; Friedman 1979; Henderson *et al.* 2003; Hock, Romanski *et al.* 1978); visual details (Mandler and Parker 1976; Friedman 1979; Pezdek *et al.* 1988, 1989); and verbally recall and recognize object descriptions (Brewer and Treyens 1981).

Incidental scene representation and memory

The evidence described in the previous sections appears to provide compelling support for the idea that detailed visual representations are generated and retained in memory during scene perception. A lingering concern from these studies, however, is the possibility that these results arise from scene processing strategies tied to the use of viewing instructions that stress scene memorization. That is, it is possible that detailed visual scene representations can be generated and retained in memory when viewers engage in intentional memory encoding, but that these representations are not typically generated incidentally during natural scene perception. If this view were correct, then the evidence for good visual memory performance obtained in prior studies might be dismissed as irrelevant to normal scene perception.

If detailed visual memory is only generated under intentional memorization instructions, then evidence for the preservation of the visual details of previously viewed objects should only be observed in intentional memorization tasks. Conversely, viewing tasks for which intentional memory encoding is unnecessary should produce no visual representation in memory. On the other hand, if detailed visual representations are typically generated and stored in memory as a natural consequence of scene perception, then evidence for the long-term preservation of visual detail should be found in both intentional and incidental memorization conditions. To investigate this issue, we have recently conducted two sets of experiments to examine the nature of the visual representations of objects generated incidentally over the course of viewing (Castelhano and Henderson 2005; Williams *et al.* 2005).

As part of his doctoral dissertation work, Carrick Williams investigated the nature of the visual memory representations that are generated for real-world objects during visual search through object arrays (Williams *et al.* 2005). Participants searched through these arrays while their eye movements were recorded. Each array contained 12 unique full-color photographs of objects from a wide variety of categories (see Fig. 9.4, top panel). In each trial, participants were asked to search for and count the number of exemplars of a specific target object, such as a green drill. Arrays contained three types of distractors: category distractors (drills that were not green), color distractors (green objects that were not drills), and unrelated distractors (objects that were neither green nor drills). After all of the arrays had been searched, participants were given a surprise forced-choice visual memory test in which they had to discriminate objects that had appeared in the arrays from memory foils that were different tokens of the same object class. Memory test items were of all three types: search targets,
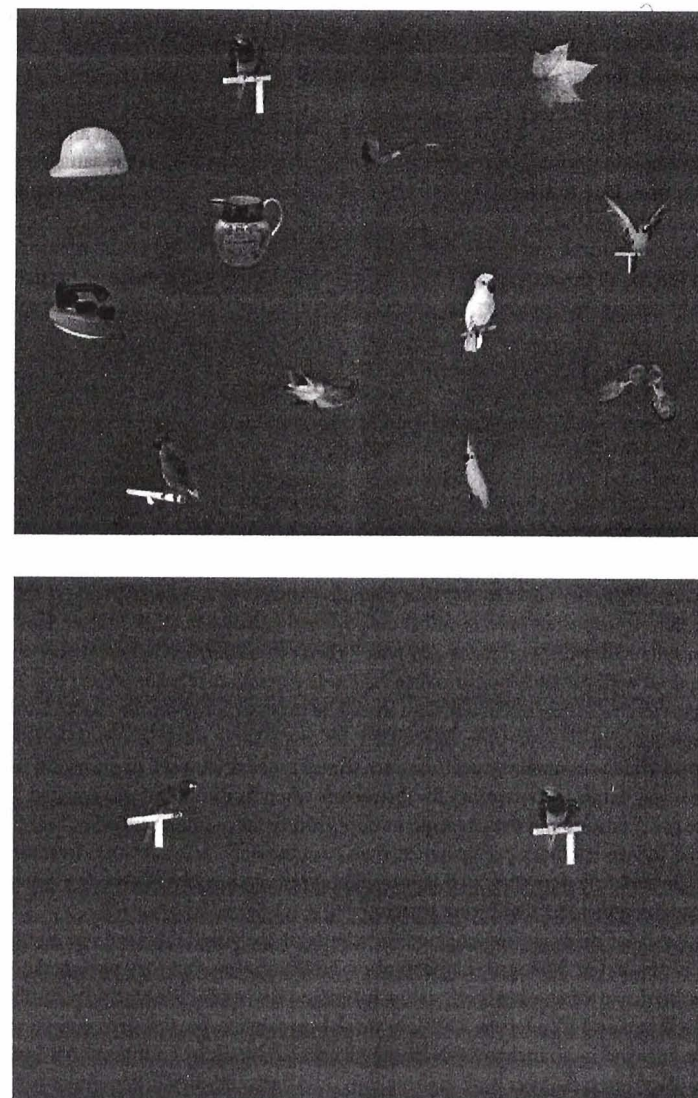

Figure 9.4 Illustration of the paradigm developed by Williams *et al.* (2005). Participants searched for a specific target (e.g. yellow bird) through object arrays containing 12 unique full-color photographs of objects (top panel). Arrays contained targets, category distractors, color distractors, and unrelated distractors. After all of the arrays had been searched, a surprise forced-choice visual memory test for all types of items was given (bottom panel). Note that the original images were presented in color.

distractors sharing either color or category with the search target, and distractors sharing neither color nor category. For example, if the test object from an array were a yellow bird, the foil would be another yellow bird (see Fig. 9.4, bottom panel). This test therefore required that relatively detailed visual information be preserved in memory. The memory task was designed to eliminate the contribution of context and semantic information to performance by presenting targets and foils that fit the same semantic description. Due to the surprise nature of the visual memory test, any learning that occurred during the search portion of the experiment was incidental.

There were three main findings in this study. First, preserved visual memory was observed for all three types of objects. This finding is remarkable because participants did not anticipate a memory test during the search task (so learning was completely incidental), and because the memory test was very stringent (test objects were presented without the context within which they were initially viewed and the foils were very similar to the targets). Second, memory was graded, with best visual memory for the search targets, intermediate memory for the related distractors, and poorest memory for the unrelated distractors. Third, this pattern was mirrored in the eye-movement data; search targets received the greatest number of fixations and the most fixation time, followed by related (color or category) distractors, followed by unrelated distractors. These last results suggest that fixation during encoding is related to the strength of the resulting memory representation. This finding is reminiscent of Friedman's (1979) observation that expected objects in scenes receive less fixation time than unexpected objects and show poorer memory when tested later. However, the results here were a bit more complex. When eye-movement behavior was directly compared with memory performance via linear regression, it became clear that search targets were remembered better than would be expected only on the basis of number of looks or total fixation time. Specifically, although there was a relationship between fixation time and memory performance for all types of objects, memory for search targets was better than memory for distractors when fixation time was equated. Thus, while eye fixations and the consequent opportunity for memory encoding was highly related to later memory performance, it was not the only factor at work. In summary, this study clearly demonstrated that visual representations for objects are generated and retained incidentally during search.

In a related study, we investigated the nature of the visual memory representation that is generated during scene perception by examining memory performance for visual information obtained either intentionally or incidentally from objects (Castelhano and Henderson 2005). In three experiments, participants viewed scenes while engaged in an incidental-learning visual search task or an intentional-learning memorization task. After both viewing tasks had been completed, a memory test for a critical object in each scene was administered, although no memory test was anticipated by the participant during the visual search task. In the memorization task, participants were instructed to view the scenes in preparation for a difficult memory test that would require knowledge of details of specific objects. In the visual search task, participants were instructed to find a specific target object in each scene, and were not told that they would receive a memory test. The top panel of Fig. 9.5 shows an example of a scene used in the experiment.
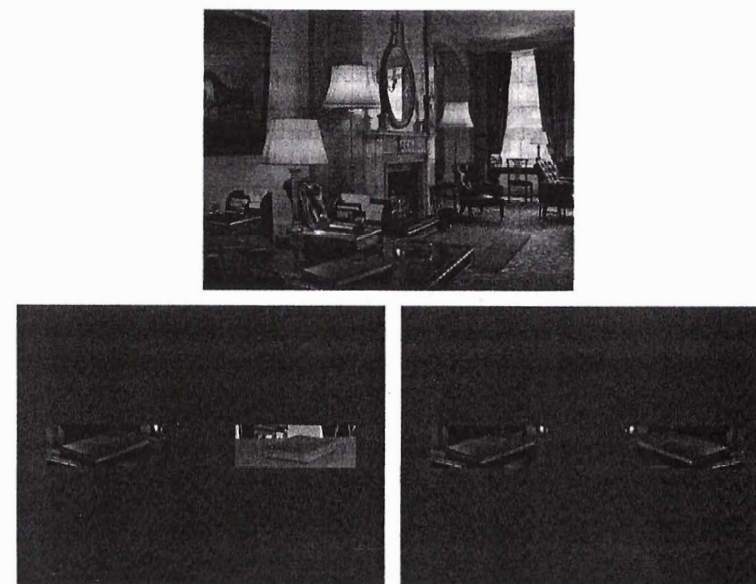
Figure 9.5 Illustration of the study reported by Castelhano and Henderson (2005). Participants searched for a specific target (e.g. ashtray) through photographs of real-world scenes (top panel). After all of the arrays had been searched, a surprise forced-choice visual memory test for the detail or the orientation of a non-target was given (bottom panels). Note that the original images were presented in color.

The test always focused on the visual properties of a specific critical object drawn from each scene. Unlike the Williams *et al.* (2005) study, the critical test objects for the search scenes were never the search targets. In the first experiment, the memory test involved discriminating between a previously seen critical object drawn from each scene and a matched foil object of the same basic-level category type (e.g. two different books, as shown in the bottom left panel of Fig. 9.5). In the second experiment, participants had to discriminate between the previously viewed orientation of the critical object and a mirror-reversed distractor version of the same object (see bottom right panel of Fig. 9.5). In both experiments, all participants took part in both the memorization and visual search tasks. In the third experiment, each participant was given only one of the two initial viewing conditions (memorization or search) from Experiment 1. This between-subjects design ensured that there was no contamination from the memorization condition to the visual search condition. The main question in the three experiments was whether long-term visual memory would be observed for objects that were incidentally encoded during scene viewing.

In all three experiments, participants showed above-chance memory for the tested objects. Furthermore, there was no evidence that memory was better in the intentional than in the incidental learning condition. As Castelhano and Henderson (2005) noted, the study involved a relatively stringent test of visual memory. Memory performance was based on a total of only 10 s of viewing time per scene and an average of <1 s of total fixation time during learning for each critical object. During the memory test, the tested objects (and their matched foils) were presented alone on a blank screen without any indication of which scene they had come from. In addition, memory performance in this study had to rely on long-term storage rather than active on-line scene representations. The retention interval varied between approximately 4–20 min between initial scene viewing and object test depending on where in the randomized sequence each scene and memory test appeared. Furthermore, the total number of objects likely to have been encoded across all of the scenes was relatively large. Using conservative estimates of object encoding (e.g. assuming that only fixated objects were encoded), Castelhano and Henderson (2005) estimated that between 373 and 440 objects were fixated and processed on average by each participant in the three experiments. All of these factors would work against finding evidence for memory of visual detail, yet such evidence was clearly obtained. Together, these results strongly suggest that visual representations are generated and stored in long-term memory as a natural consequence of scene viewing.

Scene representation, visual memory, and perceptual experience

Given the clear evidence, both historically and recently, for the creation and storage of visual object and scene representations, we might ask what leads theorists to posit the lack of such representations (e.g. O'Regan and Nöe 2001). From our perspective, this proposal has its roots in the fact that there are two traditions in vision science. The first tradition is tied to approaches that are largely concerned with attempting to explain the phenomenology of perception. Why do we experience red in the way we do? How is it that we experience a stable visual world despite the presence of saccadic eye movements? And most relevant to the topic of the current chapter, why and how do we experience a complete, detailed, full color visual world despite the fact that (a) the retinas cannot deliver this high-fidelity input within a given fixation, and (b) the visual system cannot fuse together discrete retinotopic images to generate a composite internal picture?

The second tradition, which derives from cognitive psychology and is reflected in current theoretical approaches in visual cognition as well as computer vision, is concerned with the visual representations that are available for visual and cognitive computations (and implemented in the brain in the case of human cognition) without concern for whether they give rise to perceptual experience or are open to awareness. Instead of asking what gives rise to the experience of stability across saccades (for example), those studying vision within this tradition have tended to ask about the nature of the internal representation generated across saccades, regardless of whether this representation is functional in generating experience. In the present

case, the issue from a cognitive perspective revolves around the nature of the scene representation(s) that is (potentially) generated over the course of multiple fixations and (potentially) stored in memory, again without regard for which of, or even whether, these representations give rise to perceptual experience.

In our view, a problem arises when the interpretation of data generated within the first tradition that is centered on the issue of visual experience bleeds into the second tradition, which focuses on internal representation and computation. In the case of scene perception, the problem revolves around claims about the nature of visual representation made purely on the basis of reported experience. As stated most recently and forcefully by O'Regan and Noë (2001) based on the change blindness phenomenon (though others have made similar strong statements based on change blindness in the past), 'Indeed there is no 're'-presentation of the world inside the brain ...'. (O'Regan and Nöe 2001). But we know, and have known for a very long time in cognitive psychology, that what people experience (or can report) is not necessarily a very good indication of what the brain represents. Cognitive science is rife with examples of this, but a couple of examples here should suffice to illustrate the point. First, in the study of memory, it is common place that people do not experience and cannot report memories that nonetheless exist. This sort of finding can be shown in myriad behavioral and neuro-cognitive tasks, as well as in careful assessment in the neuropsychology of amnesia. In the memory literature, the dissociation between report and representation is sometimes captured by the theoretical distinction between explicit and implicit memory. The degree to which explicit and implicit memories are supported by separate memory systems is controversial, but the dissociation between the two types of memories is not.

More directly relevant to the issue of scene representation and memory, the change blindness phenomenon similarly suggests that viewers can be unaware (or unable to report) what would otherwise appear to be salient changes to a viewed scene when those changes take place across a saccade or other visual disruption. At the same time, however, as described in an earlier section of this chapter, we have demonstrated that behavioral consequences of those changes can be observed in the absence of awareness (or at least, in the absence of report). The clearest example is increased fixation time on a changed visual region in the absence of explicit report (e.g. Henderson and Hollingworth 2003c; Hollingworth and Henderson 2002; Hollingworth et al. 2001). The increased fixation times, which can be in the order of a couple of hundred milliseconds, are themselves neither under conscious control nor consciously experienced, and they clearly indicate that there is more to an internal representation than conscious experience would lead one to believe. The implication is that one cannot draw any kind of strong conclusion about internal visual representation or computation based solely on perceptual phenomenology.

Conclusion

In this chapter we reviewed the literature concerned with the types of memory systems relevant for understanding the nature of the object and scene representations generated during scene viewing. We specifically focused on three memory epochs

important for understanding how scene representations are generated dynamically across multiple eye movements: transsaccadic memory, active on-line scene memory, and long-term scene memory. We argued that the evidence supports the conclusion that relatively detailed visual scene representations are retained over the short and long term. Furthermore, we presented recent evidence strongly suggesting that these representations are generated incidentally as a natural consequence of scene viewing. Finally, we discussed the implications of these studies and how they relate to the findings from change detection research. We conclude that the evidence strongly supports the view that relatively detailed visual representations are generated and stored in memory during active, dynamic scene perception.

Acknowledgements

References

Bahrick, H. P., and Boucher, B. (1968). Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 78, 417–422.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.

Breitmeyer, B. G., Kopft, W., and Julesz, B. (1982). The existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta Psychologica*, 52, 175–196.

Brewer, W. F., and Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207–230.

Bridgeman, B., and Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, 21, 285–286.

Bridgeman, B., and Stark, L. (1979). Omnidirectional increase in threshold for image shifts during saccadic eye movements. *Perception and Psychophysics*, 25, 241–243.

Bridgeman, B., Hendry, D., and Stark, L. (1975). Failure to detect displacements of the visual world during saccadic eye movements. *Vision Research*, 15, 719–722.

Bridgeman, B., Van der Hejiden, and Velichkovsky (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Science*, 17, 247–292.

Carlson-Radvansky, L. A. (1999). Memory for relational information across eye movements. *Perception and Psychophysics*, 61, 919–934.

Carlson-Radvansky, L. A., and Irwin, D. E. (1995). Memory for structural information across eye movements.

Castelhano, M. S., and Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition: Special Issue on Scene Perception.*

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, 27, 183–228.

Currie, C., McConkie, G., Carlson-Radvansky, L. A., and Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception and Psychophysics*, 62, 673–683.

Davidson, M. L., Fox, M. J., and Dick, A. O. (1973). Effect of eye movements on backward masking and perceived location. *Perception and Psychophyics*, 14, 110–116.

Dennett, D.C. (1991). *Consciousness explained*. Boston: Little, Brown.

Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36, 1827–1837.

Deubel, H., Schneider, W. X., and Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research*, 36, 985–996.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, 109, 75–97.

Duhamel, J. R., Colby, C. L., and Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255, 90–92.

Eriksen, C. W., and Collins, J. F. (1967). Some temporal characteristics of visual pattern recognition. *Journal of Experimental Psychology*, 74, 476–484.

Feldman, J. A. (1985). Four frames suffice: a provisional model of vision and space. *Behavioral and Brain Sciences*, 8, 265–289.

Fernandez-Duque, D., and Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, 7, 323–344.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108. 316–355.

Goodman, G. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, 12, 473–495.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In: K. Akins (ed.) *Perception: Vancouver studies in cognitive science*. Oxford: Oxford University Press, pp. 89–110.

Gysen, V., Verfaillie, K., and De Graef, P. (2002). The effect of stimulus blanking on the detection of intrasaccadic displacements of translating objects. *Vision Research*, 42, 2021–2030.

Hayhoe, M. M., Bensinger, D. G., and Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, 38, 125–137.

Henderson, J. M. (1992a). Identifying objects across eye fixations: Effects of extrafoveal preview and flanker object context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 521–530.

Henderson, J. M. (1992b). Visual attention and eye movement control during reading and picture viewing. In: K. Rayner (ed.) *Eye movements and visual cognition: scene perception and reading*. New York: Springer-Verlag, pp. 260–283.

Henderson, J. M. (1993). Visual attention and saccadic eye movements. In: G. d'Ydewalle and J. Rensbergen (ed.), *Perception and cognition: Advances in eye movement research*. Amsterdam: Netherland Science Publishers, pp. 37–50.

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, 123, 410–426.

Henderson, J. M. (1996). Visual attention and the attention-action interface. In: K. Aikens (ed.) *Perception: Vancouver studies in cognitive science*. Oxford: Oxford University Press, pp. 290–316.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science*, 8, 51–55.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, 7, 498–504.

Henderson, J. M., and Anes, M. D. (1994). Effects of object-file review and type priming on visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 826–839.

Henderson, J. M., and Ferreira, F. (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Henderson, J. M., and Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In: G. Underwood (ed.) *Eye guidance in reading and scene perception*. Oxford: Elsevier, pp. 269–283.

Henderson, J. M., and Hollingworth, A. (1999a). High-level scene perception. *Annual Review of Psychology*, 50, 243–271.

Henderson, J. M., and Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10, 438–443.

Henderson, J. M., and Hollingworth, A. (2003a). Global transsaccadic change blindness during scene perception. *Psychological Science*, 14, 493–497.

Henderson, J. M., and Hollingworth, A. (2003b). Eye movements, visual memory, and scene representation. In: M. A. Peterson and G. Rhodes (ed.), *Analytic and holistic processes in the perception of faces, objects, and scenes*. New York: Oxford University Press, pp. 356–383.

Henderson, J. M., and Hollingworth, A. (2003c). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception and Psychophysics*, 65, 58–71.

Henderson, J. M., Pollatsek, A., and Rayner, K. (1987). The effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 449–463.

Henderson, J. M., Pollatsek, A., and Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception and Psychophysics*, 45, 196–208.

Henderson, J. M., and Siefert, A. B. (1999). The influence of enantiomorphic transformation on transsaccadic object integration. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 243–255.

Henderson, J. M., and Siefert, A. B. C. (2001). Types and tokens in transsaccadic object identification: Effects of spatial position and left-right orientation. *Psychonomic Bulletin and Review*, 8, 753–760.

Henderson, J. M., Williams, C. C., Castelhano, M. S., and Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception and Psychophysics*, 65, 725–734.

Hock, H. S., Romanski, L., Galie, A., and Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory and Cognition*, 6, 423–431.

Hoffman, J. R., and Subramanian, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, 57, 787–795.

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 388–403.

Hollingworth, A., and Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition (Special Issue on Change Blindness and Visual Memory)*, 7, 213–235.

Hollingworth, A., and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.

Hollingworth, A., Schrock, G., and Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory and Cognition*, 29, 296–304.

Hollingworth, A., Williams, C. C., and Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin and Review*, 8, 761–768.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420–456.

Irwin, D. E. (1992a). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 307–317.

Irwin, D. E. (1992b). Visual memory within and across fixations. In: K. Rayner (ed.) *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag, pp. 146–165.

Irwin, D. E., and Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In: T. Inui and J. L. McClelland (ed.), *Attention and performance XVI: Information integration in perception and communication*. Cambridge, MA: MIT Press, pp. 125–155.

Irwin, D. E., and Gordon, R. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, 5, 127–155.

Irwin, D. E., Brown, J. S., and Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General*, 117, 276–287.

Irwin, D. E., Yantis, S., and Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception and Psychophysics*, 34, 49–57.

Irwin, D. E., Zacks, J.L., and Brown, J.S. (1990). Visual memory and the perception of a stable visual environment. *Perception and Psychophysics*, 47, 35–46.

Jonides, J., Irwin, D. E., and Yantis, S. (1982). Integrating visual information from successive fixations. *Science*, 215, 192–194.

Jonides, J., Irwin, D. E., and Yantis, S. (1983). Failure to integrate information from successive fixations. *Science*, 222, 188.

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897–1916.

Mack, A. (1970). An investigation of the relationship between eye and retinal image movements in the perception of movement. *Perception and Psychophysics*, 8, 291–298.

Mandler, J. M., and Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 38–48.

Mandler, J. M., and Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 386–396.

Marr, D. (1982). *Vision*. San Francisco: Freeman

McConkie, G. W. (1990). *Where vision and cognition meet*. Paper presented at the Human Frontier Science Program Workshop on Object and Scene Perception, Leuven, Belgium.

McConkie, G. W. (1991). Perceiving a stable visual world. In: J. Van Resnbergen, M. Devijver, and G. d'Ydewalle (ed.), *Proceedings of the sixth European conference on eye movements*. Leuven, Belgium: Laboratory of Experimental Psychology, pp. 5–7.

McConkie, G. W., and Currie, C. B. (1996). Visual stability while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 563–581.

McConkie, G. W., and Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In: H. Singer and R. B. Ruddell (ed.), *Theoretical models and processes in reading*. Newark, DE: International Reading Institute, pp. 137–162.

McConkie, G. W., and Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, **25**, 221–224.

Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT: Appleton-Century-Crofts.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, **19**, 155–160.

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology Special Issue: Object perception and scene analysis*, **46**, 461–488.

O'Regan, J. K., and Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, **23**, 765–768.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, **24**, 939–1031.

Parker, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 284–293.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, **9**, 441–474.

Pezdek, K., Maki, R., Valencia-Laver, D., Whetstone, T., Stockert, J. and Dougherty, T. (1988). Picture memory: Recognizing added and deleted details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 468–476.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., and Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 587–595.

Pollatsek, A., and Rayner, K. (1992). What is integrated across fixations? In: K. Rayner (ed.) *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag, pp. 166–191.

Pollatsek, A., Rayner, K., and Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, **113**, 426–442.

Pollatsek, A., Rayner, K., and Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 199–210.

Pouget, A., Fisher, S. A., and Sejnowski, T. J. (1993). Egocentric spatial representation in early vision. *Journal of Cognitive Neuroscience*, **5**, 150–161.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372–422.

Rayner, K., McConkie, G. W., and Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 529–544.

Rayner, K., and Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception and Psychophysics*, **34**, 39–48.

Rensink, R. A. (2000a). Seeing, sensing, and scrutinizing. *Vision Research*, **40**, 1469–1487.

Rensink, R. A. (2000b). The dynamic representation of scenes. *Visual Cognition Special Issue: Change blindness and visual memory*, **7**, 17–42.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, **8**, 368–373.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, **2**, 1019–1025.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition Special Issue: Change blindness and visual memory*, **7**, 377–396.

Shepard, R.N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, **6**, 156–163.

Shepherd, M., Findlay, J. M., and Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology*, **38A**, 475–491.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, **74**, 29.

Standing, L., Conezio, J., and Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, **19**, 73–74.

Trehub, A. (1977). Neuronal models for cognitive processes: Networks for learning, perception and imagination. *Journal of Theoretical Biology*, **65**, 141–169.

Verfaillie, K., De Troy, A., and Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 649–670.

Whipple, W.R., and Wallach, H. (1978). Direction-specific motion thresholds for abnormal image shifts during saccadic eye movement. *Perception and Psychophysics*, **24**, 349–355.

Williams, C. C., Henderson, J. M., and Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception and Psychophysics*.

Williams, P., and Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition Special Issue: Change blindness and visual memory*, **7**, 297–322.

Wolfe, J. M. (1999). Inattentional amnesia. In: V. Coltheart (ed.) *Fleeting memories*. Cambridge, MA: MIT Press, pp. 71–94.